

[illegible]

Background of invention

5 Field of the invention

10 [0002]

One of the techniques for sequencing nucleic acid etc. or for detecting the sequence is to utilize a DNA array. USP 5,445,934 discloses a DNA array where 100,000 or more oligonucleotide probes are bonded in 1 inch square. Such a DNA array has an advantage that many characteristics can be examined at the same time with a very small sample amount. When a fluorescence-labeled sample is poured onto such a DNA chip, DNA fragments in the sample bind to probes having a complementary sequence fixed on the DNA chip, and only that part can be discriminated by fluorescence to elucidate the sequence of the DNA fragment in the DNA sample.

Sequencing By Hybridization (SBH) is a method for examining the base sequence utilizing such a DNA array

and the details are described in USP 5,202,231. In the SBH method, all possible sequences of an oligonucleotide of a certain length are arranged on the substrate, then fully matched hybrids formed by hybridization reaction between probes and the sample DNA are detected. If a set of fully matched hybrids is obtained, the set will give an assembly of overlapping sequences with one base shift being a part of one certain sequence, of which analysis will elucidate that sequence.

[0004]

In principle, in order to examine whether a certain sequence is present or not in a DNA specimen, hybridization reaction is carried out with a probe having a complementary sequence, and the presence or absence of hybridization is detected. In practice, however, it is very difficult to judge the presence or absence of one sequence by using one complementary probe and hybridization, because even when fully matched hybrids are compared, the fluorescence intensities of the hybrids differ from each other according to their sequence. In particular, GC content in the sequence greatly affects the stability of the hybrid. Further, sequences not exactly complementary but containing one base mismatch also form a hybrid to emit fluorescence. Such a hybrid has lower stability and weaker fluorescence compared with a fully matched hybrid of the same sequence, but it is often observed

that such a mismatch hybrid has a stronger fluorescence than a full-matched hybrid of different sequence. In addition, the stability of one mismatch hybrid greatly varies according to the location of the mismatch in the hybrid. When the mismatch locates at the terminus, a relatively stable hybrid is obtained. When the mismatch locates at the center of the hybrid, the hybrid becomes unstable because the consecutiveness of the complementary strand is broken. Thus, at present, various factors are complexedly participating in the hybrid stability, and the absolute value (standard value) for the fluorescence intensity to judge whether the hybrid is full matched or not is not obtained. Also, conditions for detecting the fluorescence solely from the full matched hybrid, eliminating fluorescence from one-base mismatched hybrids have not been determined.

[0005]

In order to dissolve the difference of the hybrid stability due to the sequence, a method using tetramethylammonium chloride is described in Proc. Natl. Acad. Sci. USA Vol. 82, pp.1585-1588 (1985). However, the above described problems have not been solved perfectly.

[0006]

A method for judging whether a hybrid being a perfect match is described in Science vol. 274 p.610-

614, 1996, in which a 15-mer oligonucleotide probe and 15-mer oligonucleotides having the same sequence except for one mismatching base at the center of the sequence are prepared, and the fluorescence intensity of the hybrid with the probe (perfect match) is compared with those of hybrids with other one-base mismatching oligonucleotides, and only when the intensity of the perfect match is stronger, it is judged positive.

[0007]

Based on the method above, USP 5,733,729 discloses a method using a computer for more accurate calling, where the fluorescence intensities of the hybrids are compared by using a computer to know the base sequence of a sample.

[0008]

In these methods, it is necessary to locate the subject nucleotide to be examined in the center of a probe and to prepare a set of four probes each having one of four bases at the position. It is also necessary to prepare such a probe set for each of overlapping sequences with one base shift. As described above, they use 15-mer oligonucleotides and determine the perfect match by comparing with other three types of probes having one-base mismatch at the center. It is said that more accuracy can be obtained by evaluating the stability of the hybrids theoretically or empirically. In addition, if the base length of the region to be

examined is L, the number of probes will be $4 \times L$ (e.g., 20 probes for 5 bases).

[0009]

Although the above described methods using
5 mismatches are excellent in that the call is relatively
easy by comparing with one-base mismatches at the same
position of the same sequence and that the number of
probes may be small (in SBH, 1024 types of probes are
required for the similar analyses), they have
10 significant defects that accurate information can not
be obtained when there are two base mismatches in the
same region or when there is a base deletion or
insertion.

[0010]

15 On the other hand, the SBH method may solve the
above described problems and in principle, it may cope
with any variation. Call, however, is rather difficult,
since intensity of one-base mismatch in one sequence is
stronger than that of a full match in another sequence
20 and since stability of the hybrid differs considerably
according to the position of the mismatch in the
sequence even if it is an one-base mismatch. As a
result, a full match, one-base and two-base mismatches
(continuous or discontinuous) can not be simply called
25 from the fluorescence intensities. Accordingly complex
analyses including theoretical predictions, comparison
between individual sequences and accumulation of

empirical parameters are required.

[0011]

Furthermore, in order to determine the sequence of a gene by reading fluorescence intensities of hybrids for each probe followed by data analysis, a large-scale computer system as well as a detector for reading arrays is required. This is a big obstacle in the way of simple gene diagnosis using the DNA array.

[0012]

10 SUMMARY OF THE INVENTION

In view of such problems, the present invention provides a method of accurate gene sequencing not requiring complex analyses.

[0013]

15 As described above, the fluorescence intensity of a hybrid is controlled by various factors. Thus, when a probe having about 12 mer to 25 mer in length is used, it is hard to exclude the fluorescence due to hybrids having one-base mismatch. On the other hand, it is
20 relatively easy to obtain the conditions for inhibiting formation of two-base mismatch hybrids regardless of position, continuity or discontinuity of the two-base mismatch, when a probe of 12 mer to 25 mer in length is used.

25 [0014]

The present invention has been achieved based on such a finding characterized in that spots of mismatch

hybrids containing a predetermined number of mismatches are taken into account as well as a spot of a perfect match hybrid.

[0015]

5 According to one embodiment of the present invention, there is provided a method for identifying an unknown base sequence present in a target single-stranded nucleic acid comprising the steps of:

10 (a) preparing a probe array in which single-stranded nucleic acid probes of No. 1 to No. n ($n \geq 2$) are arranged as isolated spots on a substrate, the probes each having a base sequence complementary to one of plural base sequences expected to be the unknown base sequence;

15 (b) reacting a single-stranded nucleic acid, which has a base sequence fully complementary to a base sequence of one of the single-stranded nucleic acid probes and is fluorescence-labeled, with the probe array under such conditions that single-stranded
20 nucleic acids complementary to each other form a double-stranded nucleic acid;

 removing the unreacted labeled single-stranded nucleic acid, and

 measuring fluorescence intensity of each spot of
25 the probe array to obtain a first template pattern showing a relationship between location of the probes and fluorescent characteristics;

(c) performing the same operation as the step (b) for each of remaining single-stranded nucleic acid probes using a second to a nth single-stranded nucleic acid, and obtaining template patterns of No. 2 to No. n showing a relationship between location and fluorescent characteristics of the probes;

(d) performing the same operation as the step (b) using a sample containing the target single-stranded nucleic acid of unknown base sequence to obtain a sample pattern showing relationship between a position and fluorescent characteristics; and

(e) comparing the sample pattern obtained in the step (d) with n pieces of template patterns obtained in the steps (b) and (c), to identify a template pattern showing substantially the same pattern as the sample pattern and identifying the base sequence of the single-stranded nucleic acid used for the preparation of the identified template pattern as the unknown base sequence of the target single-stranded nucleic acid.

[0016]

According to another embodiment of the present invention, there is provided a method for identifying an unknown base sequence present in a target single-stranded nucleic acid comprising the steps of:

(a) preparing a probe array in which single-stranded nucleic acid probes of No. 1 to No. n ($n \geq 2$) are arranged as isolated spots on a substrate, the

probes each having a base sequence complementary to one of plural base sequences expected to be the unknown base sequence;

(b) reacting a single-stranded nucleic acid which
5 has a base sequence fully complementary to a base sequence of one of the single-stranded nucleic acid probes and is fluorescence-labeled, with the probe array under such conditions that single-stranded nucleic acids complementary to each other form a
10 double-stranded nucleic acid;

removing the unreacted labeled single-stranded nucleic acid, and

measuring fluorescence intensity of each spot of the probe array to obtain a first template pattern
15 showing a relationship between location of the probes and fluorescent characteristics;

(c) analyzing the first template pattern to locate probes and to calculate a mean value of fluorescence intensities (F_i) of the double-stranded nucleic acids
20 having i of mismatched base pairs, where i is an integer not less than 1;

(d) calculating a difference ($F_1, 0$) between the fluorescence intensity of the fully complementary double-stranded nucleic acid without mismatch (F_0) and
25 the mean value of the fluorescence intensities of the double-stranded nucleic acids having one-base mismatch (F_1), further calculating a difference (F_{i+1}, i)

between a fluorescence intensity of a double-stranded nucleic acid having $(i+1)$ base mismatches (F_{i+1}) and a fluorescence intensity of a double-stranded nucleic acid having i -base mismatches (F_i) , and identifying i being F_{i+1} , $i \ll F_i$, $i-1$;

(e) assuming a target DNA which base sequence is complementary to the second probe sequence, then obtaining the second template pattern formed by the probe position where the number of mismatched base pairs to the target having the complementary sequence to the second probe sequence is not more than i ;

(f) performing the same operation as the step (e) for each of remaining single-stranded nucleic acid probes using a third to a n th single-stranded nucleic acid, and obtaining template patterns of No. 3 to No. n showing a relationship between location and fluorescent characteristics of the probes, wherein the template patterns are formed from the positions of the probes having a base sequence that forms mismatched base pairs in a number not more than i ;

(g) performing the same operation as the step (b) using a sample containing the target single-stranded nucleic acid of unknown base sequence to obtain a sample pattern showing relationship between a position and fluorescent characteristics; and

(h) comparing the sample pattern obtained in the step (g) with n pieces of template patterns obtained in

the steps (b), (c) and (e), to identify a template pattern showing essentially the same pattern as the sample pattern and identifying the base sequence of the single-stranded nucleic acid used for the preparation of the identified template pattern as the unknown base sequence of the target single-stranded nucleic acid.

[0017]

According to the present invention, patterns of positive spots on the substrate are taken as images, and, the unknown sequence can be analyzed by comparing the images with the predicted pattern to identify the unknown genetic sequence easily.

[0018]

Hybridization conditions which allows complete discrimination between one-base mismatch and two-base mismatch are also disclosed.

[0019]

BRIEF DESCRIPTION OF THE DRAWINGS

FIG. 1 shows a pattern of an arrangement when 64 types of probes are used;

FIG. 2 shows a pattern of the arrangement showing positive spots formed with a target nucleic acid;

FIG. 3 shows patterns of the arrangement showing positive spots formed with variant sequences of the target nucleic acid;

FIG. 4 shows is a pattern obtained in Example 1 with fluorescence intensities.

FIG. 5 is an expected pattern in Example 2;

FIG. 6 is a pattern obtained in Example 2 with a fluorescence threshold of 10%; and

FIG. 7 is a pattern obtained in Example 3 with
5 fluorescence intensities.

DETAILED DESCRIPTION OF THE INVENTION

The present invention is explained in detail.

Call using fluorescence image

10 One embodiment of the present invention particularly effective when bases which may cause mismatching exist close to each other. Herein, this will be explained using 5'GATGGGNCTCNGTTCAT3' as an example, this sequence includes a base sequence
15 corresponding to the 248th and 249th amino acids (hereinafter AA248 and AA249) of tumor suppressor gene p53. This example is only to explain this invention roughly, not to limit the present invention to a specific array form or probe arrangement. The concept
20 of the present invention that the result is treated as an image is applicable to any form of arrays. The SBH method is naturally subjected to the analysis of the present invention.

[0020]

25 In the above example, when a full set of probes is prepared by replacing the base represented by N with any of four bases (A, G, C, T), that is, when three

bases (no need for continuity) are examined, $4^3 = 64$ probes are arranged on the substrate. $4^5 = 1024$ probes are required to examine five bases.

[0021]

- 5 An example of the arrangement when 64 types of probes are used is shown in FIG. 1.

[0022]

- 10 In this example, in the upper left quarter of the array of 64 probes, are arranged the probes of which first N is A (probe number: 1-16), while in the lower left quarter, the probes of which first N is G (probe number: 17-32). Similarly, in the upper right quarter, probes of which first N is C (probe number: 33-48) are arranged and in the lower right, those having the first
- 15 N of T (probe number: 49-64). In each region, the probes having the second N of A are positioned in the first column from the left, G, C and T for the second, third and fourth columns respectively. Also, probes having the third N of A are positioned in the first row
- 20 from the top in each region, G, C and T in the second, third and fourth rows respectively. As a result, for example, the sequence of 5'GATGGGACTCAAGTTCAT3' corresponds to the upper left corner spot. A target nucleic acid being 5'ATGAACCGGAGGCCCATC3', which
- 25 corresponds to the normal gene, is expected to form a hybrid with a probe DNA 5'GATGGGCCTCCGGTTCAT3' which is positioned at the crosspoint of the third column from

the right and the third row from the top.

[0023]

Now the case where one-base mismatches are included in a template pattern for determining the sequence of the gene will be explained. In this case, if the fully matching sequence is the probe 42 (normal), one-base mismatching sequences to be called positive correspond to 9 points (shadowed circles), forming a pattern together with the perfect match point as shown in FIG. 2.

[0024]

On the other hand, the pattern change is observed with a target nucleic acid having a variant sequence to be identified, as shown in FIG. 3.

[0025]

In the present invention, images of the expected fluorescent patterns composed of such full match and one-base mismatch hybrids are input into a computer memory device or the like in beforehand, and the call is performed by comparing the fluorescent image obtained by a predetermined method with the memory. Herein, detailed quantitative data of the fluorescence intensity of positive spots are not required. Simple judgement on whether the fluorescence is stronger than the threshold value which has been determined experimentally enables simple and automatic calling using a computer etc.

[0026]

Setting of threshold

When a probe of about 18 mer is used, the threshold is preferably set between the fluorescence intensity of the one-base mismatch and that of the two-base mismatch. Although the fluorescence intensity depends on the sequence or the reaction conditions, 50% to 25%, more preferably 30% to 20%, of the highest fluorescent intensity (normally of the full match hybrid) may be used as the threshold. When the length of the probe is shorter, the threshold will be lower.

[0027]

Fluorescence of those having three-base mismatch will be below 10% of the maximum fluorescence, allowing complete discrimination.

[0028]

FIG. 4 shows the spots which fluoresce at an intensity higher than 10% of the maximum fluorescence corresponding to the full match and one-base mismatch hybrids.

[0029]

A more specific calling method will be described with the above example.

[0030]

When the hybridization reaction is carried out very selectively, strong fluorescence appears only at one point (the full match). When the sensitivity is

increased gradually or the stringency in reaction conditions is reduced, as expected from FIG. 3 in the above arranged example, the one-base mismatch points will appear in a row and a column crossing at the full match point. However, the actual fluorescent image is not always such that three spots each align in a row and column around a strong fluorescent point. Since six points not always have similar fluorescence intensity due to the hybrid stability difference, not all of the spots can be detected. However, at least some spots would be seen on those lines. At the same time, the remaining one-base mismatches may fluoresce at the expected positions, although the intensity might be weaker than other spots.

15 [0031]

Sometimes, the full match hybrid and one-base mismatch hybrids may have similar fluorescence intensity to give a pattern consisting of the expected 10 spots of the full match and one-base mismatches.

20 [0032]

Although the fluorescence intensity of two-base mismatch hybrids sometimes exceed the threshold, they can be distinguished easily because of the divergence from the expected pattern.

25 [0033]

Thus, the method of the present invention where calling is performed by comparing the expected pattern

with the actually obtained fluorescent image has a feature that the presence or absence of variation in the test gene can be easily determined and, at the same time, the nature of the variation (which base(s) is changed to what base(s)) can be determined.

[0034]

Further, when the result of hybridization using 64 probes is assessed, the idea of pattern assessment has an advantage that calling is more reliable than with only one spot. Since the hybrids with 64 DNA probes differ in heat stability between individual sequences, it is not ensured that the full match hybrid is always far more stable and radiates stronger fluorescence. In addition, it is often impossible to determine the strongest and full match spot due to the foreign matters on the substrate or the artifacts during the hybridization reaction. In this point, calling by pattern can compensate certain variation of fluorescence intensity, if any.

[0035]

Probe length

The probe length used for the present invention is approximately 8 mer to 30 mer, more preferably 12 mer to 25 mer. When it is shorter than 8 mer, stability of the hybrids having one-base mismatch is low and the fluorescence from the full match is superior, while when it is longer than 30 mer, the fluorescence of two-

base mismatches sometimes is (for example, when mismatches locate at the both ends) stronger than that of one-base mismatches.

[0036]

5 Conditions of hybridization reaction

Preferable hybridization conditions are as follows: A substrate is soaked completely in a sample solution and heated for heat-denaturing both the DNA probes on the substrate and the sample DNA, then the substrate and the solution are cooled slowly to perform the hybridization reaction. The salt concentration of the reaction mixture without formamide is desirably below 100 mM.

[0037]

15 Appropriate temperature for heat denaturation is 60°C or higher, preferably 80°C or higher. The temperature for heat denaturation is determined depending on the stability of the substrate itself, length and concentration of the test DNA, type of the labeling compound. For example, with such a substrate prepared by binding DNA to a resin layer formed on the surface of the substrate, sometimes the resin layer is destroyed by heating at a high temperature. On the other hand, substrates prepared using a silane coupling agent are rather heat-stable and can be heated to a higher temperature. When the test DNA is a single-stranded DNA, the intramolecular double-stranded

structure is melt at 70°C or more, while when the sample is a double-stranded DNA or long single-stranded DNA, it is necessary to melt the double-stranded structure by heating at a higher temperature or by adding a denaturing agent such as formamide. Time required for heat denaturation is 10 min or more, depending on the microassay size and the volume of the sample solution.

[0038]

10 The hybridization conditions are determined according to the conventional method where temperature and salt concentration are changed considering the length and sequence of the probes, and the type of the test sample. The suitable conditions for discriminating
15 extremely similar sequences as in the present invention are 45°C for over 3 hours in a solution containing 100 mM of sodium chloride. However, as the reaction time is greatly affected by the sample concentration, it is not limited to the above reaction conditions. With a
20 sample of high concentration, calling within 3 hours is possible, while with a dilute sample, 10 hours or more of the reaction time is required. When formamide is added, the concentration of sodium chloride should be increased.

25 Preparation of DNA array

 How to prepare the DNA array suitable for the hybridization reaction of the present invention is

exemplified below. However, since the purpose of the present invention is to provide a simple method for evaluating the hybridization pattern on the substrate to determine the base sequence of a sample, the substrate preparation method is not specifically limited.

[0039]

DNA probes may be covalently bonded to the substrate by reacting the probes with functional groups on the substrate. The following is a method of coupling reaction between a maleimide group on the glass surface with an SH group at the end of DNA.

[0040]

Maleimide groups can be incorporated onto the surface of a substrate, first, by introducing amino groups with an amino silane coupler onto the substrate, and then reacting the amino groups with a reagent containing N-(6-maleimidocaproyloxy)succinimide (EMCS reagent: Dojin Co., Ltd.). Introduction of an SH group to DNA can be performed by use of 5'-Thiol-Modifier C6 (Glen Research Company) on a DNA-automatic synthesizer.

[0041]

Spots of the DNA probes are formed by the ink jet method on the substrate, then the probe DNA is fixed by the reaction between the maleimide groups on the substrate and the SH groups at the end of the DNA.

[0042]

A DNA solution suitable for ink jet ejection to the maleimide-substrate is one containing glycerin, urea, thiodiglycol or ethylene glycol, acetylenol EH (Kawaken Fine Chemical Company-made) and isopropyl alcohol. Particularly, a solution containing 7.5% of glycerin, 7.5% of urea, 7.5% of thiodiglycol and 1% of acetylenol EH is preferable.

[0043]

The array substrate to which DNA has been bonded is then soaked in an aqueous solution of 2% bovine serum albumin for 2 hours for blocking. Now it is ready for a hybridization reaction.

[0044]

Examples

The invention will be described in the following Examples in more detail.

Example 1: pattern recognition I

1. Probe Design

It is well known that in the base sequence CGGAGG corresponding to the AA248 and AA249 of the tumor suppressor gene p53, frequently observed variation is those the first C to T, the second A to G for AA248, and the third G to T for AA249. Accordingly, aiming at these three positions, 64 types of probes were designed.

[0045]

That is, the designed nucleic acid are 18-mer
nucleic acids harboring variegated above mentioned six
bases sandwiched between the common sequences, to be
represented by 5'ATGAACNNGAGNCCCATC3' where N
5 corresponds to any of 4 bases, A, G, C and T. Actual
probes to detect the above sequence should be have a
complementary sequence of 5'GATGGGNCNNGTTCAT3'.
[0046]

FIG. 1 shows an arrangement of 64 types of DNA
10 probes on a substrate. Each sequence (SEQ ID NOs: 1 to
64) is specifically shown in Table 1.

004660-962460

[0047]

[Table 1]

Sequence Number (SEQ ID NO:)	Sequence	Sequence Number (SEQ ID NO:)	Sequence
1	GATGGGACTCAAGTTCAT	33	GATGGGCCTCAAGTTCAT
2	GATGGGACTCAGGTTTCAT	34	GATGGGCCTCAGGTTTCAT
3	GATGGGACTCACGTTTCAT	35	GATGGGCCTCACGTTTCAT
4	GATGGGACTCATGTTTCAT	36	GATGGGCCTCATGTTTCAT
5	GATGGGACTCGAGTTCAT	37	GATGGGCCTCGAGTTCAT
6	GATGGGACTCGGGTTCAT	38	GATGGGCCTCGGGTTCAT
7	GATGGGACTCGCGTTCAT	39	GATGGGCCTCGCGTTCAT
8	GATGGGACTCGTGTTTCAT	40	GATGGGCCTCGTGTTTCAT
9	GATGGGACTCCAGTTCAT	41	GATGGGCCTCCAGTTCAT
10	GATGGGACTCCGGTTCAT	42	GATGGGCCTCCGGTTCAT
11	GATGGGACTCCCGTTCAT	43	GATGGGCCTCCCGTTCAT
12	GATGGGACTCCTGTTTCAT	44	GATGGGCCTCCTGTTTCAT
13	GATGGGACTCTAGTTCAT	45	GATGGGCCTCTAGTTCAT
14	GATGGGACTCTGGTTCAT	46	GATGGGCCTCTGGTTCAT
15	GATGGGACTCTCGTTCAT	47	GATGGGCCTCTCGTTCAT
16	GATGGGACTCTTGTTTCAT	48	GATGGGCCTCTTGTTTCAT
17	GATGGGGCTCAAGTTCAT	49	GATGGGTCTCAAGTTCAT
18	GATGGGGCTCAGGTTTCAT	50	GATGGGTCTCAGGTTTCAT
19	GATGGGGCTCACGTTTCAT	51	GATGGGTCTCACGTTTCAT
20	GATGGGGCTCATGTTTCAT	52	GATGGGTCTCATGTTTCAT
21	GATGGGGCTCGAGTTCAT	53	GATGGGTCTCGAGTTCAT
22	GATGGGGCTCGGGTTCAT	54	GATGGGTCTCGGGTTCAT
23	GATGGGGCTCGCGTTCAT	55	GATGGGTCTCGCGTTCAT
24	GATGGGGCTCGTGTTTCAT	56	GATGGGTCTCGTGTTTCAT
25	GATGGGGCTCCAGTTCAT	57	GATGGGTCTCCAGTTCAT
26	GATGGGGCTCCGGTTCAT	58	GATGGGTCTCCGGTTCAT
27	GATGGGGCTCCCGTTCAT	59	GATGGGTCTCCCGTTCAT
28	GATGGGGCTCCTGTTTCAT	60	GATGGGTCTCCTGTTTCAT
29	GATGGGGCTCTAGTTCAT	61	GATGGGTCTCTAGTTCAT
30	GATGGGGCTCTGGTTCAT	62	GATGGGTCTCTGGTTCAT
31	GATGGGGCTCTCGTTCAT	63	GATGGGTCTCTCGTTCAT
32	GATGGGGCTCTTGTTTCAT	64	GATGGGTCTCTTGTTTCAT

5' ATGAACCGGAGGCCCATC3' which is the sequence
corresponding to the normal gene is expected to form a

hybrid with the DNA probe 42 of 5'GATGGGCCTCCGGTTCAT3'
located at the third point from the right and from the
top.

[0048]

5 In experiment of 64 hybrid formation, fluorescence
from the one-base mismatch hybrids is also expected in
addition to that from the full match hybrid. an
expected pattern of the fluorescence from the full
match hybrid and one-base mismatch hybrids is shown in
10 FIG. 2.

[0049]

2. Preparation of substrate introduced with maleimide
group

Substrate Cleaning

15 A glass plate of 1 inch square was placed in a
rack and soaked in an ultrasonic cleaning detergent
overnight. Then, after 20 min of ultrasonic cleaning,
the detergent was removed by washing with water. After
rinsing the plate with distilled water, ultrasonic
20 treatment was repeated in a container filled with
distilled water, for additional 20 min. Then the plate
was soaked in a prewarmed 1N sodium hydroxide solution
for 10 min, washed with water and then distilled water.

[0050]

25 Surface treatment

Then the plate was soaked in an aqueous solution
of 1% silane coupling agent (product of Shin-Etsu

Chemical Industry: Trade name KBM 603) at a room temperature for 20 min, thereafter nitrogen gas was blown on the both sides blowing off water to dryness. The silane coupling treatment was completed by baking the plate in an oven at 120°C for 1 hour. Subsequently, 2.7 mg of EMCS (N-(6-maleimidocaproyloxy) succinimide: Dojin Company) was weighed and dissolved in a 1 : 1 solution of DMSO/ethanol (final concentration: 0.3 mg/ml). The glass substrate treated with the silane coupling agent was soaked in this EMCS solution for 2 hours to react the amino group of the silane coupling agent with the succimide group of EMCS. At this stage, the maleimide group of EMCS is transferred to the glass surface. After that, the glass plate was washed with ethanol, and dried with nitrogen gas to be used for a coupling reaction with DNA.

3. Coupling of DNA to the substrate

Synthesis of 64 DNA probes

The above 64 types of probe DNAs each having an SH group (thiol group) at the 5' terminus were synthesized by Becks Co., Ltd. at our request.

Ejection of DNA probes

The above 64 types of DNAs were ejected respectively as follows. Each DNA was dissolved in water and diluted with SG Clear (aqueous solution containing 7.5% of glycerin, 7.5% of urea, 7.5% of

thiodiglycol and 1% of acetylenol EH), to a final concentration of 8 μ M. Then 100 μ l of this DNA solution was filled into a nozzle of a BJ printer Head BC 62 (Canon) modified to eject a small amount, and to eject six solutions per head. Two heads were used at a time so that 12 types of DNAs could be ejected at once, and the heads were changed 6 times so that 64 spots of 64 types of DNAs were formed on the glass plates independently.

10 [0051]

Sixty four probes were spotted with a diameter of 70 μ m and a pitch of 200 μ m to form a matrix of 8 \times 8. After that, the plate was left standing in a humidified chamber for 30 min for linking reaction of the probe DNA to the substrate.

15

Hybridization reaction

- Blocking reaction

After completion of the reaction, the substrate was washed with a 1 M NaCl/50 mM phosphate buffer solution (pH 7.0) to wash out thoroughly the DNA solution on the glass surface. Then, this was soaked in an aqueous solution of 2% bovine serum albumin and allowed to stand for 2 hours to carry out a blocking reaction.

20

- Preparation of model sample DNA

Rhodamine labeled DNA No. 1 (SEQ ID NO: 65) of the same length as the probes but having the normal sequence of p53 gene was prepared. The sequence is shown below and rhodamine is bonded to the 5' terminus.

[0052]

No. 1 : 5'Rho-ATGAACCGGAGGCCCATC3'

- Hybridization conditions

Two milliliters of a 10 nM model sample DNA solution containing 100 mM NaCl was applied to the DNA array substrate in a hybridization bag, and the bag was initially heated at 80°C for 10 min. Then the temperature of the incubator was lowered to 45°C and the reaction was continued for 15 hours.

5. Detection

- Detection method

The detection was performed by connecting an image analysis processing apparatus, ARGUS (a product of Hamamatsu Photonics) to a fluorescence microscope (a product of Nikon).

- Result

The fluorescence intensities obtained from the model hybridization reaction with the labeled DNA No. 1 (18-mer) are shown in FIG. 4. The maximum value of the fluorescence intensity was obtained at the spot of probe 42 which is fully complementary to DNA No. 1. Taking this intensity as the maximum value (1.0), the

threshold is set at 10% of this value and the spots having higher intensity are painted dark.

[0053]

The spots of probes 10, 26, 41, 46 and 58 of one-
base mismatch hybrids have fluorescence higher than the
threshold, and it is understood that the location is
well coincident with FIG. 2 of the expected pattern.
By lowering the threshold further, in addition to the
above 5 spots, the spots of other one-base mismatch
probes appeared around the full matching probe in
vertical and horizontal lines, in coincidence with the
expected pattern.

[0054]

Example 2: Pattern recognition II

A DNA array of 64 types of probes was prepared in
the same manner as in Example 1, and the hybridization
reaction was performed using a rhodamine-labeled DNA
No. 2 as a model sample. The DNA No. 2 (SEQ ID NO: 66)
has a sequence complementary to the No. 46 probe of
FIG. 1.

[0055]

No. 2: 5'Rho-ATGAACCAGAGGCCCATC3'

The reaction conditions of hybridization are the
same as in Example 1.

[0056]

FIG. 5 is an expected pattern consisting of the
perfect match and one-base mismatch hybrids, and the

resulted pattern obtained as in Example 1 is shown in FIG. 6. The threshold is set at 10% of the maximum value. When the detected spots are painted dark, the result is well corresponding to the expectation.

5 [0057]

Example 3: Pattern recognition III

An experiment was carried out in the same manner as in Example 2 except that the concentration of the sample DNA used for the hybridization reaction was 5 nM and the reaction was carried out at 40°C overnight. The result obtained is shown in FIG. 7.

[0058]

If the threshold is set as 50%, fluorescence was detected at the positions (shaded parts) of Nos. 34 and 62 probes (one-base mismatch) in addition to No.46 (full match), and with further reduction of the threshold to 10%, the result was coincident with the expected pattern. In this case, Nos. 6, 22 and 54 of two-base mismatch probes were detected, but the two-base mismatch can be distinguished from the one-base mismatch as the deviation from the expected pattern of one-base mismatch, and No. 46 can be called as the full matched probe.

SEQUENCE LISTING

<110>Canon INC.

<120>Method of analyzing base sequence of nucleic acid

<130>CFO 15718

<150>JP 263506/2000

<160>66

<210>1

<211>18

<212>DNA

<213>Artificial sequence

<220>

<223>Sample oligonucleotide

<400>1

gatgggactc aagttcat

<210>2

<211>18

<212>DNA

<213>Artificial sequence

<220>

<223>Sample oligonucleotide

<400>2

gatgggactc aggttcat

<210>3

<211>18

<212>DNA

<213>Artificial sequence

<220>

401630-363450

<223>Sample oligonucleotide

<400>3

gatgggactc acgttcat

<210>4

<211>18

<212>DNA

<213>Artificial sequence

<220>

<223>Sample oligonucleotide

<400>4

gatgggactc atgttcat

<210>5

<211>18

<212>DNA

<213>Artificial sequence

<220>

<223>Sample oligonucleotide

<400>5

gatgggactc gagttcat

<210>6

<211>18

<212>DNA

<213>Artificial sequence

<220>

<223>Sample oligonucleotide

<400>6

gatgggactcgggtcat

0044303460

<210>7

<211>18

<212>DNA

<213>Artificial sequence

<220>

<223>Sample oligonucleotide

<400>7

gatgggactc gcgttcat

<210>8

<211>18

<212>DNA

<213>Artificial sequence

<220>

<223>Sample oligonucleotide

<400>8

gatgggactc gtgttcat

<210>9

<211>18

<212>DNA

<213>Artificial sequence

<220>

<223>Sample oligonucleotide

<400>9

gatgggactc cagttcat

<210>10

<211>18

<212>DNA

441630:324460

<213>Artificial sequence

<220>

<223>Sample oligonucleotide

<400>10

gatgggactc cggttcat

<210>11

<211>18

<212>DNA

<213>Artificial sequence

<220>

<223>Sample oligonucleotide

<400>11

gatgggactc cggttcat

<210>12

<211>18

<212>DNA

<213>Artificial sequence

<220>

<223>Sample oligonucleotide

<400>12

gatgggactc ctgttcat

<210>13

<211>18

<212>DNA

<213>Artificial sequence

<220>

<223>Sample oligonucleotide

0044634 = 963460

<400>13

gatgggactc tagttcat

<210>14

<211>18

<212>DNA

<213>Artificial sequence

<220>

<223>Sample oligonucleotide

<400>14

gatgggactc tggttcat

<210>15

<211>18

<212>DNA

<213>Artificial sequence

<220>

<223>Sample oligonucleotide

<400>15

gatgggactc tggttcat

<210>16

<211>18

<212>DNA

<213>Artificial sequence

<220>

<223>Sample oligonucleotide

<400>16

gatgggactc ttgttcat

<210>17

<211>18

<212>DNA

<213>Artificial sequence

<220>

<223>Sample oligonucleotide

<400>17

gatggggctc aagttcat

<210>18

<211>18

<212>DNA

<213>Artificial sequence

<220>

<223>Sample oligonucleotide

<400>18

gatggggctc aggttcat

<210>19

<211>18

<212>DNA

<213>Artificial sequence

<220>

<223>Sample oligonucleotide

<400>19

gatggggctc acgttcat

<210>20

<211>18

<212>DNA

<213>Artificial sequence

File: 96040600

<220>

<223>Sample oligonucleotide

<400>20

gatggggctc atgttcat

<210>21

<211>18

<212>DNA

<213>Artificial sequence

<220>

<223>Sample oligonucleotide

<400>21

gatggggctc gagttcat

<210>22

<211>18

<212>DNA

<213>Artificial sequence

<220>

<223>Sample oligonucleotide

<400>22

gatggggctc gggttcat

<210>23

<211>18

<212>DNA

<213>Artificial sequence

<220>

<223>Sample oligonucleotide

<400>23

404630 = 96344600

gatggggctc gcgttcat

<210>24

<211>18

<212>DNA

<213>Artificial sequence

<220>

<223>Sample oligonucleotide

<400>24

gatggggctc gtgttcat

<210>25

<211>18

<212>DNA

<213>Artificial sequence

<220>

<223>Sample oligonucleotide

<400>25

gatggggctc cagttcat

<210>26

<211>18

<212>DNA

<213>Artificial sequence

<220>

<223>Sample oligonucleotide

<400>26

gatggggctc cggttcat

<210>27

<211>18

00436-0341

<212>DNA

<213>Artificial sequence

<220>

<223>Sample oligonucleotide

<400>27

gatggggctc ccgttcat

<210>28

<211>18

<212>DNA

<213>Artificial sequence

<220>

<223>Sample oligonucleotide

<400>28

gatggggctc ctgttcat

<210>29

<211>18

<212>DNA

<213>Artificial sequence

<220>

<223>Sample oligonucleotide

<400>29

gatggggctc tagttcat

<210>30

<211>18

<212>DNA

<213>Artificial sequence

<220>

40430 = 30460

<223>Sample oligonucleotide

<400>30

gatggggctc tggttcat

<210>31

<211>18

<212>DNA

<213>Artificial sequence

<220>

<223>Sample oligonucleotide

<400>31

gatggggctc tcgttcat

<210>32

<211>18

<212>DNA

<213>Artificial sequence

<220>

<223>Sample oligonucleotide

<400>32

gatggggctc ttgttcat

<210>33

<211>18

<212>DNA

<213>Artificial sequence

<220>

<223>Sample oligonucleotide

<400>33

gatgggcctc aagttcat

101830-934460

<210>34

<211>18

<212>DNA

<213>Artificial sequence

<220>

<223>Sample oligonucleotide

<400>34

gatgggcctc aggttcat

<210>35

<211>18

<212>DNA

<213>Artificial sequence

<220>

<223>Sample oligonucleotide

<400>35

gatgggcctc acgttcat

<210>36

<211>18

<212>DNA

<213>Artificial sequence

<220>

<223>Sample oligonucleotide

<400>36

gatgggcctc atgttcat

<210>37

<211>18

<212>DNA

141834 = 3333333333

<223>Sample oligonucleotide

<400>40

gatgggcctc gtgttcat

<210>41

<211>18

<212>DNA

<213>Artificial sequence

<220>

<223>Sample oligonucleotide

<400>41

gatgggcctc cagttcat

<210>42

<211>18

<212>DNA

<213>Artificial sequence

<220>

<223>Sample oligonucleotide

<400>42

gatgggcctc cggttcat

<210>43

<211>18

<212>DNA

<213>Artificial sequence

<220>

<223>Sample oligonucleotide

<400>43

gatgggcctc ccgttcat

<210>44

141430 9694960

<211>18

<212>DNA

<213>Artificial sequence

<220>

<223>Sample oligonucleotide

<400>44

gatgggcctc ctgttcat

<210>45

<211>18

<212>DNA

<213>Artificial sequence

<220>

<223>Sample oligonucleotide

<400>45

gatgggcctc tagttcat

<210>46

<211>18

<212>DNA

<213>Artificial sequence

<220>

<223>Sample oligonucleotide

<400>46

gatgggcctc tggttcat

<210>47

<211>18

<212>DNA

<213>Artificial sequence

bioRxiv preprint doi: <https://doi.org/10.1101/000000>; this version posted March 1, 2014. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

<220>

<223>Sample oligonucleotide

<400>47

gatgggcctc tcgttcat

<210>48

<211>18

<212>DNA

<213>Artificial sequence

$\langle 220 \rangle$

<223>Sample oligonucleotide

<400>48

gatgggcctc ttgttcat

<210>49

<211>18

<212>DNA

<213>Artificial sequence

<220>

<223>Sample oligonucleotide

<400>49

gatgggtctc aagttcat

<210>50

<211>18

<212>DNA

<213>Artificial sequence

 $\langle 220 \rangle$

<223>Sample oligonucleotide

<400>50

gatgggtctc aggttcat

<210>51

<211>18

<212>DNA

<213>Artificial sequence

<220>

<223>Sample oligonucleotide

<400>51

gatgggtctc acgttcat

<210>52

<211>18

<212>DNA

<213>Artificial sequence

<220>

<223>Sample oligonucleotide

<400>52

gatgggtctc atgttcat

<210>53

<211>18

<212>DNA

<213>Artificial sequence

<220>

<223>Sample oligonucleotide

<400>53

gatgggtctc gagttcat

<210>54

<211>18

40430-364460

<213>Artificial sequence

<223>Sample oligonucleotide

gatgggtctc gggttcat

<211>18

<213>Artificial sequence

<223>Sample oligonucleotide

gatgggtctc gcgttcat

<211>18

<213>Artificial sequence

<223>Sample oligonucleotide

gatgggtctc gtgttcat

<211>18

<213>Artificial sequence

<220>

<223>Sample oligonucleotide

<400>57

gatgggtctc cagttcat

<210>58

<211>18

<212>DNA

<213>Artificial sequence

<220>

<223>Sample oligonucleotide

<400>58

gatgggtctc cggttcat

<210>59

<211>18

<212>DNA

<213>Artificial sequence

<220>

<223>Sample oligonucleotide

<400>59

gatgggtctc ccgttcat

<210>60

<211>18

<212>DNA

<213>Artificial sequence

<220>

<223>Sample oligonucleotide

<400>60

gatgggtctc ctgttcat

101530-9696960

<211>18

<213>Artificial sequence

 $\langle 220 \rangle$

<223>Sample oligonucleotide

<400>61

gatgggtctc tagttcat

<210>62

<211>18

<212>DNA

<213>Artificial sequence

 $\langle 220 \rangle$

<223>Sample oligonucleotide

<400>62

gatgggtctc tggttcat

<210>63

<211>18

<212>DNA

<213>Artificial sequence

 $\langle 220 \rangle$

<223>Sample oligonucleotide

<400>63

gatgggtctc tcgttcat

<210>64

<211>18

<212>DNA

<220>

<400>64

<210>65

<212>DNA

 $\langle 220 \rangle$ $\langle 400 \rangle_{65}$

<210>66

<212>DNA

$\langle 220 \rangle$

<400>66

atgaaccaga ggcccatc